<u>Mesure de dispersion</u>

Définitions :

<u>Moyenne</u> $\overline{X}$

C'est la somme des résultats divisé par le nombre de résultats.

Pour la moyenne d'un échantillon, on utilise le symbole $\overline{X}$ (se nomme « X barre).

Écart-type (σ pour une population **ou s** pour un échantillon**)**

L'écart-type se rapproche de l'écart moyen et donne une bonne indication sur la dispersion des données à la moyenne. Elle est égale à la racine carrée de la variance.

n = taille de la population ou de l'échantillon

<u>Avec données non-groupées</u>

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{n}} \quad \text{ou} \quad s = \sqrt{\frac{\sum (x_i - \overline{X})^2}{n-1}}$$

<u>Avec données dans un tableau de fréquence (effectif)</u>

$$\sigma = \sqrt{\frac{\sum f_i (x_i - \mu)^2}{n}} \quad \text{ou} \quad s = \sqrt{\frac{\sum f_i (x_i - \overline{X})^2}{n-1}}$$

<u>Avec données dans un tableau avec des classes (point milieu)</u>

$$\sigma = \sqrt{\frac{\sum f_i (m_i - \mu)^2}{n}} \quad \text{ou} \quad s = \sqrt{\frac{\sum f_i (m_i - \overline{X})^2}{n-1}}$$

Exemple : reprenons la distribution précédente

46 57 64 70 70 76 84 93 $\qquad \overline{X} = 70$

$$s^2 = \frac{(46-70)^2 + (57-70)^2 + (64-70)^2 + (70-70)^2 + (70-70)^2 + (76-70)^2 + (84-70)^2 + (93-70)^2}{7}$$

$$s^2 = \frac{1542}{7} = 220{,}29$$

$$s = \sqrt{220{,}29} = \mathbf{14{,}84}$$

Si on considère cette distribution comme une population, on va calculer l'écart-type avec sigma.

$$\sigma = \sqrt{\frac{1542}{8}} = \sqrt{192{,}75} = \mathbf{13{,}88}$$

On remplace n par n-1 dans un échantillon pour obtenir un écart-type plus près de l'écart moyen en général. La différence entre σ et s est négligeable lorsque n > 50.

Un petit écart-type signifie que les données sont concentrées ou peu étalées de chaque côté de la moyenne (plus homogène). Un grand écart-type signifie le contraire.

Analysons le pourcentage de données compris entre différents intervalles.

$\overline{X} = 70$ et $s = 14{,}84$

Intervalle $[\overline{X}-s, \overline{X}+s]$ = [55,16, 84,84] = 6/8 = 75% des données de la distribution se retrouvent dans cet intervalle.

Intervalle $[\overline{X}-2s, \overline{X}+2s]$ = [40,32, 99,68] = 8/8 = 100% des données de la distribution se retrouvent dans cet intervalle.

Les données se retrouvent habituellement entre $[\overline{X}-3s, \overline{X}+3s]$